

빅데이터 연구 논문의 주제 분야 연관관계 분석: 동시 인용 관계를 적용하여*

Subject Association Analysis of Big Data Studies: Using Co-citation Networks

곽철완 (Chul-Wan Kwak)**

초 록

본 연구의 목적은 빅데이터 연구 논문의 주제 분야 간의 연관관계를 분석하는데 있다. 동시 인용 관계를 적용하여 분석 대상의 주제 분야를 추출하였으며, R 프로그램의 Apriori 알고리즘을 이용하여 연관관계의 규칙을 분석하고, arulesViz 패키지를 사용하여 시각화하였다. 연구 결과 22개 주제 분야가 추출되었는데, 이들 주제 분야는 3가지 군집으로 구분되었다. 주제 분야의 연관관계 유형을 분석한 결과, 연관관계의 복잡성에 따라 '전문형', '일반형', '확대형'으로 구분되었다. 전문형에는 문헌정보학, 신문방송학 등이 포함되었고, 일반형에는 정치외교학, 무역학, 관광학 등이 포함되었고, 확대형에는 기타인문학, 사회과학일반, 관광학일반 등이 포함되었다. 이 연관관계는 빅데이터 연구자가 한 주제 분야를 인용할 때 관계가 있는 다른 주제 분야를 인용하는 경향을 보여주는 것으로, 도서관에서 학술정보서비스를 위해 연관관계를 활용한 서비스를 고려해야 할 필요가 있다.

ABSTRACT

The purpose of this study is to analyze the association among the subject areas of big data research papers. The subject group of the units of analysis was extracted by applying co-citation networks, and the rules of association were analyzed using Apriori algorithm of R program, and visualized using the arulesViz package of R program. As a result of the study, 22 subject areas were extracted and these subjects were divided into three clusters. As a result of analyzing the association type of the subject, it was classified into 'professional type', 'general type', 'expanded type' depending on the complexity of association. The professional type included library and information science and journalism. The general type included politics & diplomacy, trade, and tourism. The expanded types included other humanities, general social sciences, and general tourism. This association networks show a tendency to cite other subject areas that are relevant when citing a subject field, and the library should consider services that use the association for academic information services.

키워드: 빅데이터, 연관관계 분석, 동시 인용 관계, R 프로그램, 학술정보서비스
big data, association analysis, co-citation networks, R program,
academic information services

* 이 논문은 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2013S1A5A2A01017315).

** 강남대학교 산업데이터사이언스학부 교수(ckwak@kangnam.ac.kr)

■ 논문접수일자: 2018년 2월 11일 ■ 최초심사일자: 2018년 3월 8일 ■ 게재확정일자: 2018년 3월 21일
■ 정보관리학회지, 35(1), 13-32, 2018. [http://dx.doi.org/10.3743/KOSIM.2018.35.1.013]

1. 서론

1.1 연구의 필요성 및 목적

최근 우리나라에서는 제4차 산업혁명시대라는 이름으로 ‘빅데이터’가 사회적으로 중요한 이슈가 되고 있다. 이와 관련하여 문헌정보학을 비롯하여, 통계학, 산업공학, 경영학, 관광학 등 다양한 학문 분야에서 빅데이터 연구가 진행되고 있으며, 연구 범위는 급속도로 확산되고 있다. 빅데이터 주제는 특정 학문의 연구 분야가 아니라 다양한 학문에서 다양한 형태로 연구되고 있는 융복합적인 연구 분야이다. 그 범위는 빅데이터의 본질적인 특성을 밝히려는 연구에서부터 빅데이터를 응용하는 연구까지 매우 광범위하다. 예를 들면, 문헌정보학의 이정미(2013)는 빅데이터의 개념과 도서관에서 활용에 대해 연구하였으며, 컴퓨터학의 백봉현과 하일규(2017)는 빅데이터의 저장과 시각화에 대한 연구를 하였고, 경영과학의 이환수, 임동원, 조항정(2013)은 빅데이터 시대에 개인정보 과잉에 대해 연구하였다. 이러한 빅데이터 연구 분야의 다양성은 도서관 입장에서 본다면 기존의 학술정보서비스와 다른 관점의 서비스 필요성을 제기하고 있다.

오늘날까지 대부분의 학술정보서비스는 동일한 학문 분야 내에서 이루어졌다. 예를 들면 문헌정보학 연구자는 문헌정보학 분야의 연구 정보를 조사하면 대부분 문제를 해결할 수 있었다. 도서관에서도 학술정보서비스는 해당 학문 분야에 대한 연구 정보를 제공하는데 초점을 맞추었다. 하지만, 빅데이터와 같은 새로운 연구 주제의 출현은 기존의 학술정보서비스 방

법으로는 이용자의 요구를 만족시키기 충분하지 못하다. 연구자들이 다양한 학문 분야의 연구 정보를 조사하여 적합한 정보를 자신의 연구에 활용하고 있는 상황에서, 학술정보서비스 역시 이에 적합한 연구 정보를 제공해주어야 할 것이다. 예를 들면, 문헌정보학에서 빅데이터에 관한 연구가 통계학 혹은 신문방송학의 빅데이터 연구와 연관관계를 가지고 있다면, 도서관에서도 문헌정보학의 빅데이터 연구자에게 통계학 혹은 신문방송학에서 연구된 빅데이터 연구 정보를 제공해 준다면 많은 도움이 될 것이다. 이러한 학술정보서비스 확대를 위해 빅데이터 연구에서 상호 연관관계를 가지고 있는 주제 분야를 파악할 필요가 있다. 이에 본 연구의 목적은 빅데이터 연구 논문의 주제 분야 간에 연관관계를 분석하는데 있다.

1.2 선행 연구

특정 연구 분야의 동향 혹은 특징을 분석하기 위해 학술지에 발표된 연구 논문의 분석은 분석 대상을 기준으로 크게 세 가지로 구분된다. 가장 일반적인 분석의 대상은 논문의 주제이며, 다음으로 논문의 저자, 그리고 주제와 저자가 동시에 포함되는 경우로 구분된다. 첫째, 논문 주제에 대한 분석은 해당 학문 분야의 연구 동향 혹은 발전과정 연구에 많이 사용되었다. 논문의 주제 파악을 위해 주로 논문 제목, 초록, 키워드 등이 많이 활용되었다. 이재운(2015)은 문헌정보학 분야의 핵심 주제를 파악하기 위해 KCI에 수록된 논문 중 10회 이상 인용된 159개 논문을 분석의 대상으로 삼았다. 이들 논문의 동시 인용 빈도를 통하여 군집을 도출하여, 논문의

제목 기준으로 동시인용 네트워크를 구축하였다. 논문 제목으로부터 연구자가 주제 분야를 도출하여 문헌정보학 분야에서 연구가 활발하게 진행되는 주제 분야를 분석하였다. 정동열(2016)은 국내 문헌정보학 분야 연구 논문의 주제를 제목과 초록으로부터 추출하여 연도별로 분석하였다. 논문 주제 분야를 크게 도서관학, 정보학, 서지학, 기록관리학으로 구분한 후, 이를 기준으로 10년 단위로 묶어서 우리나라 문헌정보학의 발전 및 변천과정을 큰 틀에서 기술하였다. 이 연구는 문헌정보학 분야 논문 전체를 조사한 것으로 큰 의미가 있다. 이보람과 정은경(2016)은 Web of Science에 수록된 학술지를 통해 빅데이터 주제와 관련된 주제 분야의 학제적 구조를 파악하였다. 연구 결과 빅데이터를 연구하는 중심 학문 영역은 컴퓨터공학으로 정보시스템, 인공지능, 소프트웨어 등이 포함되어 있고, 과학기술, 커뮤니케이션, 법학 등에서 연구가 진행되고 있다고 분석하였다.

주제 추출을 위해 단어의 출현 빈도를 이용한 연구도 있다. 이기현, 정효정, 송민(2015)은 국외 문헌정보학 분야의 학술지에 게재된 논문 서지사항을 대상으로 단어 출현 빈도를 이용하여 논문의 주제와 연구 방법을 분석하였다. 세부적인 분석을 위해 '가중 주제-방법 네트워크 군집 분석'을 적용하였고, 결과로 최근 문헌정보학 주제로 네트워크 분석과 텍스트 마이닝 군집이 형성되었다고 분석하였다. 조재인(2011)은 문헌정보학 분야 연구의 주제 분포와 주제의 관계성을 파악하기 위해 논문 제목의 단어 출현 빈도를 사용하였다. 분석을 위해 네트워크 텍스트 분석을 적용하였으며, 시각화를 위해 NodeXL을 사용하였다. 김희정(2011)은 국외

웹 아카이빙에 관련된 연구의 논문 제목, 초록, 키워드를 대상으로 동시출현단어 네트워크를 이용하여 주제 영역을 분석하였다.

주제 분석을 위해 논문의 키워드를 활용한 연구도 있다. 장임숙, 장덕현, 이수상(2011)은 다문화 분야 연구에서 저자가 부여한 키워드를 대상으로 동시단어 네트워크와 k-core 분석을 적용하였다. 이와 유사하게 최이랑(2015)은 국내 기록관리학 분야 연구 논문의 키워드를 네트워크 텍스트 분석을 통하여 주제 관계를 분석하였다. 타 학문 분야인 기술경영 분야에서 고재창, 조근태, 조운호(2013)는 해외 논문에 수록된 키워드를 수집하여 키워드 네트워크 분석을 통하여 연구 동향을 분석하였으며, 조수곤과 김성범(2012)은 산업공학 분야의 논문 초록의 키워드를 연관 규칙과 사회연결망 분석을 통하여 연구 동향을 분석하였다.

둘째, 논문의 저자를 분석의 대상으로 삼은 많은 연구는 특정 분야의 연구자 네트워크 분석을 목적으로 하였다. 박선영과 정은경(2012)은 국외 경제학 분야 연구 논문을 대상으로 하여 피인용빈도가 높은 저자를 분석의 대상으로 삼았다. 인용빈도 부여 방식을 4가지 구분한 후, 복수저자기반의 저자동시인용분석을 통하여 경제학 분야의 지적구조를 파악하고, 인용빈도 부여 방법을 비교하였다. 이수상(2013)은 기록관리학 분야의 투고 논문과 인용 논문의 저자를 대상으로 하여, 공저 네트워크, 저자동시인용 네트워크, 저자서지결합 네트워크의 중심성과 연구성과의 연관성을 분석하였다.

셋째, 논문의 주제와 저자를 동시에 분석의 대상으로 삼은 연구가 있다. 손혜인과 남영준(2016)은 기록관리학 분야 연구의 공동연구

자 소속 기관과 주제를 분석하였다. 공동연구자 소속 기관의 네트워크와 소속 기관과 주제의 네트워크를 제시하였다. 박옥남(2011)은 KCI에 등재된 학술지에서 정보조직 분야 논문을 대상으로 하여, 키워드, 저자 소속기관, 저자를 네트워크 텍스트 분석을 통하여 관계를 분석하였다.

국외 연구로 Mishra, Luo, Jiang, Papadopoulos, Dubey(2017)는 빅데이터 분야 연구를 대상으로 동시인용분석을 이용하여 2011-15년 사이에 발표된 57편의 연구를 분석하였다. Peng, Zhang, Zhong, Zhu(2012)은 2000-09년 사이에 발표된 인터넷 관련 논문을 대상으로 저자, 제목, 잡지명, 초록, 저자 키워드, 인용 문헌 등에서 주제를 추출하여 군집을 만든 후, 단어 동시출현 네트워크 분석을 통하여 핵심 연구 영역을 분석하였다. Zhao와 Strotmann(2008, 2014)은 1996-2005, 2006-10년 두 기간 동안의 정보학의 핵심 연구 분야를 서지결합법을 통하여 분석하였다.

2. 데이터 수집 및 분석 방법

2.1 데이터 수집

국내 빅데이터 분야의 주제 분야 연관성을 분석하기 위해 2017년 12월 현재 한국연구재단의 KCI(한국학술지인용색인)에서 제공하고 있는 논문을 기본 데이터로 선택하였다. KCI의 논문검색창에 '빅데이터'를 입력한 결과 1,774편의 논문이 검색되었으며, 이 결과물 중 피인용횟수가 10회 이상인 논문 60편을 선택하였다. 선택된 논문을 수록하고 있는 학술지를 대상으로 KCI에서 부여한 주제 분야를 파악한 결과 총 22가지였다. 이중 산업공학과 학제간연구 분야가 각각 7편으로 가장 많았고, 다음으로 기타과학기술학과 법학이 각각 5편 순이었다(〈표 1〉참조). 이들 논문들은 2012년에서 2016년 사이에 발행되었는데, 연도별로 살펴보면 2012년에 6편, 2013년에 24편, 2014년에 21편, 2015년에 8편, 2016년에 1편이 발행되었다.

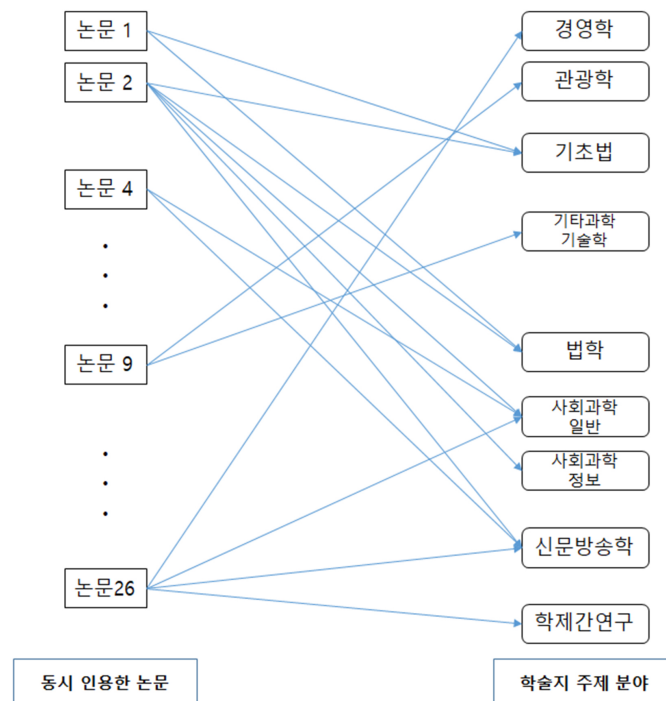
〈표 1〉 분석대상 논문을 수록하고 있는 학술지의 주제 분야와 논문 수

주제 분야	논문 수	주제 분야	논문 수
경영학	3	법학	5
관광학	4	사회과학	2
관광학일반	1	사회과학일반	4
기초법	1	사회과학정보	2
기타공학일반	2	산업공학	7
기타과학기술학	5	신문방송학	3
기타법학	3	예술일반	1
기타인문학	1	정보통신이론	1
디자인	1	정치외교학	1
무역학	1	통계학	4
문헌정보학	1	학제간연구	7

2.2 주제 분야 범주화

주제 분야 연관성 분석을 위해 피인용횟수가 높은 논문에서 동시 인용 관계를 추출하였다. 빅데이터에 관한 피인용횟수가 10회 이상인 60편의 논문을 인용한 논문은 총 1,029편(중복 포함)이었으며, 이 중 2편 이상의 논문을 동시 인용한 논문은 총 134편(중복 제외)이었다. 134편의 논문 중에는 적게는 2편의 논문을 동시 인용한 논문으로부터 13편의 논문을 동시 인용한 논문까지 매우 다양하게 분포되어 있었다. 이들 134편의 논문이 동시 인용한 60편의 논문에 관한 주제 분야는, 60편의 논문을 수록한 학술지에 대해 KCI에서 부여한 주제 분야를 사용하였다. 학술지의 주제 분야를 논문의 주제 분야

로 사용한 이유는, 연구자가 논문 발표를 위해 선택한 학술지는 논문 주제와 밀접한 관련이 있다고 가정하였기 때문이다. 그러므로 본 연구의 분석에서는 학술지의 주제 분야와 논문의 주제 분야를 동일하게 취급하였다. 논문 주제 분야에 대한 범주 작성 방법에 대해 예를 들면, <그림 1>에서 [논문 1]은 여러 논문을 인용하였으나, 그 논문 중 피인용횟수가 높은 60편에 포함된 논문은 2편이며, 이 논문이 수록된 학술지의 주제 분야가 '기초법'과 '법학'이므로 {'기초법', '법학'}으로 범주를 구성하였다. 이러한 방법으로 134편의 논문과 피인용횟수가 높은 60편 논문을 수록한 학술지의 주제 분야에 대한 매트릭스를 작성하였다(<부록 1> 참조).



<그림 1> 주제 분야 범주화 방법

2.3 분석방법

빅데이터 분야 논문의 주제 분야 연관성 분석을 위해 R 프로그램의 Apriori 알고리즘을 이용하여 연관관계의 규칙을 분석하였다. Apriori 알고리즘은 연관규칙 분석을 위해 사용되는데, 사건 A가 일어났을 때, 사건 B가 일어날 확률을 나타내는 것으로 사건 A와 사건 B 사이에 연관관계에 대한 규칙을 찾는 것을 목표로 한다(김용, 2012). Apriori 알고리즘은 각 주제 분야들의 연관관계에 대한 규칙을 제시한다. 예를 들면, 다음 장에 있는 <표 2>에서 보여주는 것과 같이 특정 연구자가 빅데이터 연구에서 ‘사회과학일반’ 논문을 인용한 경우, ‘신문방송

학’ 논문을 인용할 신뢰도가 57.9%임을 나타내고 있다. 또한 주제 분야 사이에 관계를 시각화하기 위해 R 프로그램의 arulesViz 패키지를 사용하였다.

3. 데이터 분석

3.1 주제 분야 연관관계 모델링

빅데이터 연구의 주제 분야 22가지의 연관관계 조사를 위해 Apriori 알고리즘(supp=0.0048¹⁾, conf=0.5)을 통해 조사한 결과, 749가지 연관관계가 도출되었는데, <표 2>는 support(지지

<표 2> 주제 분야 연관관계 모델링 및 결과

	lhs	=>	rhs	support*	confidence**	lift	count
[1]	{기초법}	=>	{법학}	0.09701493	0.8125000	3.754310	13
[2]	{기타법학}	=>	{법학}	0.08208955	0.7857143	3.630542	11
[3]	{사회과학일반}	=>	{신문방송학}	0.08208955	0.5789474	3.878947	11
[4]	{신문방송학}	=>	{사회과학일반}	0.08208955	0.5500000	3.878947	11
[5]	{무역학}	=>	{관광학}	0.06716418	0.9000000	4.466667	9
[6]	{기타공학일반}	=>	{산업공학}	0.05223881	0.6363636	1.894949	7
[7]	{기타법학}	=>	{기초법}	0.05223881	0.5000000	4.187500	7
[8]	{문헌정보학}	=>	{산업공학}	0.04477612	0.5454545	1.624242	6
[9]	{기초법, 기타법학}	=>	{법학}	0.03731343	0.7142857	3.300493	5
[10]	{관광학일반}	=>	{관광학}	0.02985075	0.8000000	3.970370	4
[11]	{관광학일반}	=>	{기타인문학}	0.02238806	0.6000000	3.654545	3
[12]	{관광학, 관광학일반}	=>	{기타인문학}	0.02238806	0.7500000	4.568182	3
[13]	{관광학일반, 기타인문학}	=>	{관광학}	0.02238806	1.0000000	4.962963	3
[14]	{관광학, 사회학}	=>	{기타인문학}	0.02238806	1.0000000	6.090909	3
[15]	{기타인문학, 사회과학}	=>	{관광학}	0.02238806	1.0000000	4.962963	3
[16]	{기타인문학, 무역학}	=>	{관광학}	0.02238806	1.0000000	3.722222	3
중략							
[75]	{정치외교학}	=>	{디자인}	0.00746269	1.0000000	44.666667	1
[76]	{정치외교학}	=>	{통계학}	0.00746269	1.0000000	7.052632	1
[77]	{정치외교학}	=>	{기타인문학}	0.00746269	1.0000000	6.090909	1
[78]	{정치외교학}	=>	{산업공학}	0.00746269	1.0000000	2.977778	1
	<이하 생략>		* 지지도 = 두 주제 분야가 동시에 인용될 횟수 / 전체 인용 횟수 ** 신뢰도 = 두 주제 분야가 동시에 인용될 횟수 / 한 주제 분야의 논문이 인용될 횟수				

1) 5회 미만의 관찰은 과대 해석할 위험이 있으므로 이를 최소화하기 위해 $5/1029=0.0048$ 지지도를 적용함.

도)가 높은 연관관계 사례 일부를 제시하였다. 결과를 살펴보면, 특정 연구자가 '기초법'과 '법학' 논문을 동시에 인용할 지지도는 전체 논문 인용 횟수의 9.7%로 나타났으며, '기초법' 논문을 인용한 연구자가 '법학' 논문을 동시에 인용할 확률인 신뢰도는 81.3%로 나타났다. 이 결과에서 지지도가 10% 미만으로 나타난 것은 전체 주제 분야의 수가 22가지로 많기 때문으로 판단된다. 하지만 '기초법' 논문을 인용한 연구자가 '법학' 논문을 인용할 신뢰도가 81.3%로 높기 때문에 주제 분야 사이의 연관관계가 매우 높다고 볼 수 있다. 그 밖의 주제 분야 연관관계에 대해 특정 연구자가 '문헌정보학'과 '산업공학' 논문을 동시에 인용할 지지도는 4.5%이지만, '문헌정보학' 논문을 인용한 연구자가 '산업공학' 논문을 인용할 신뢰도는 54.5%를 차지하였다. 또한, '기타인문학' 및 '사회과학' 논문을 동시에 인용한 연구자가 '관광학' 논

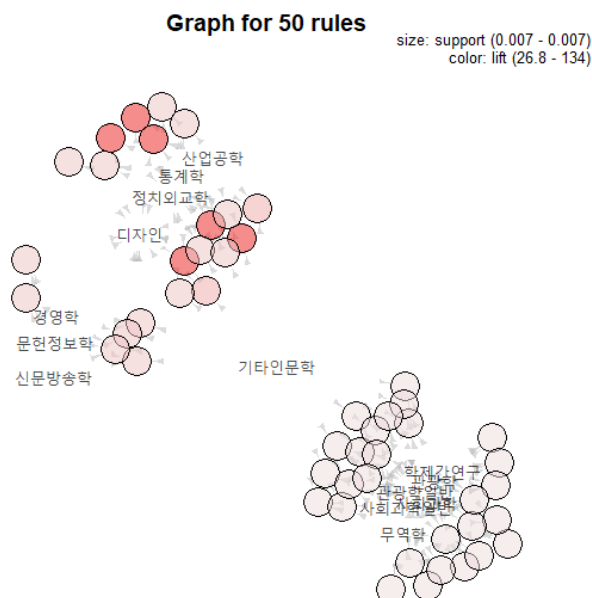
문을 인용할 신뢰도는 100%를 보여주고 있어, 이들 주제 분야 사이에 연관관계는 매우 높다고 판단된다.

3.2 주제 분야 연관관계 시각화

3.2.1 전주제 분야 연관관계

22가지 주제 분야의 연관관계를 R 프로그램의 arulesViz 패키지를 사용하여 시각화하면, 연관관계가 더욱 명확하게 나타난다. 22가지 주제 분야는 크게 3가지 군집으로 구분되는데, 군집 1은 소규모로 {문헌정보학, 경영학, 신문방송학}이 여기에 포함된다. 군집 2는 {통계학, 산업공학, 정치외교학, 디자인} 등이 포함되어 있다. 군집 3은 {사회과학, 사회과학일반, 관광학, 관광학일반, 무역학, 학제간연구} 등으로 구성되어 있다(〈그림 2〉 참조).

주제 분야 사이의 연관관계를 연결선을 이용



〈그림 2〉 주제 분야 연관관계 그래프

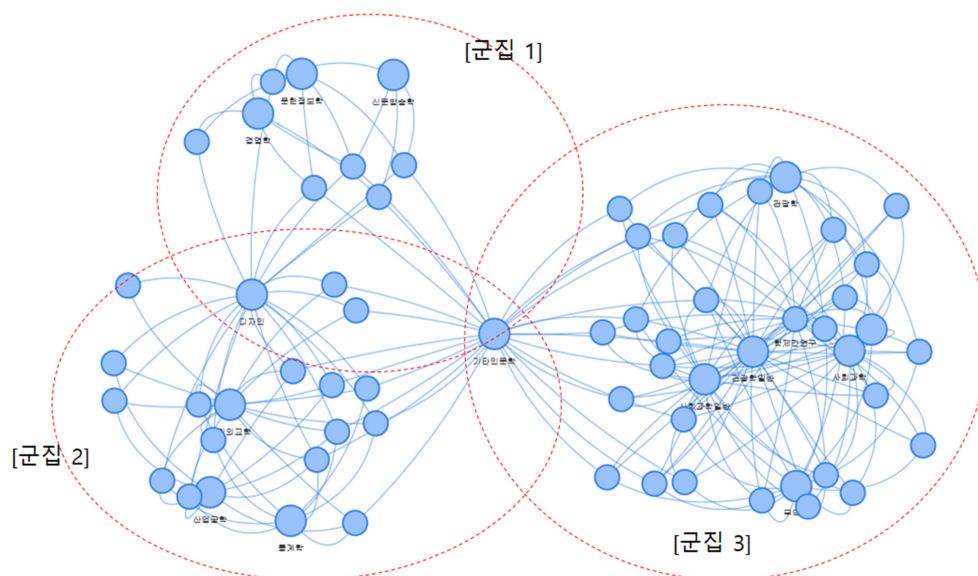
해서 살펴보면 더욱 명확해진다. <그림 3>에서 보듯이 빅데이터 연구의 주제 분야는 크게 3가지 군집으로 나뉘며, 3가지 군집을 연결하는 주제 분야가 '기타인문학'이며, 군집 1과 군집 2를 연결하는 주제 분야는 '디자인'임을 나타내고 있다. 문헌정보학이 포함된 군집 1은 군집 가운데에 특정 주제 분야가 없이 분산되어 있지만, 군집 2에서는 '정치외교학'이 가운데 위치하며, 군집의 외곽에 '통계학', '산업공학'과 '디자인'이 위치해 있다.

'정치외교학'처럼 군집의 가운데에 위치해 있다는 것은 특정 연구자가 빅데이터 연구에서 정치외교학 논문을 인용하면서 인근에 위치한 빅데이터에 관한 여러 주제 분야의 논문을 동시에 인용하는 경향이 있다는 것을 나타낸다(<그림 3> 참조). 예를 들면, 특정 연구자가 빅데이터에 관한 정치외교학 논문을 인용할 때, 빅데이터에 관한 '디자인', '통계학', '산업공학' 논문

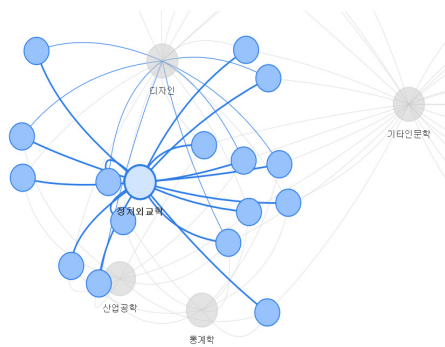
을 각각 인용할 신뢰도가 100%임을 보여주고 있다(<표 2> 참조). 군집 3에서도 '관광학일반'이 중심부에 위치하고 있고, '사회과학일반', '무역학', '사회과학', '학제간연구', '관광학'이 위치해 있다. 여기서도 빅데이터에 관한 '관광학일반'을 인용하는 논문은 주변에 위치한 여러 주제의 논문을 동시에 각각 인용하는 경향이 있다는 것을 보여주고 있다(<그림 4> 참조).

3.2.2 주제 분야 연관관계 유형

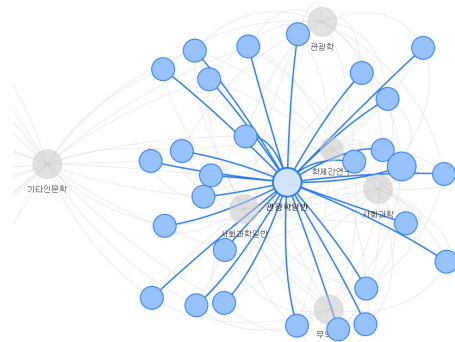
개별 주제 분야의 연관관계는 직접적으로 주제 분야 노드 사이에 연결되어 있는 것이 아니라, 중간 연결 노드를 통해서 연관관계를 보여준다. 중간 연결 노드를 통한 연관관계는 R 프로그램에서 해당 연결 노드를 클릭하여 확인할 수 있다. 연관관계 유형은 3가지 연관관계를 가지고 있는 것에서 28가지를 가지고 있는 것까지 그 유형이 매우 다양하였다. 연관관계 수를 기준으



<그림 3> 주제 분야 연관관계 그래프(연결선 포함)



[정치외교학 연관관계]



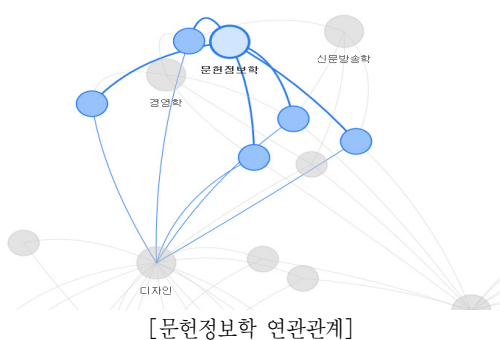
[관광학일반 연관관계]

〈그림 4〉 정치외교학 및 관광학일반 연관관계 사례

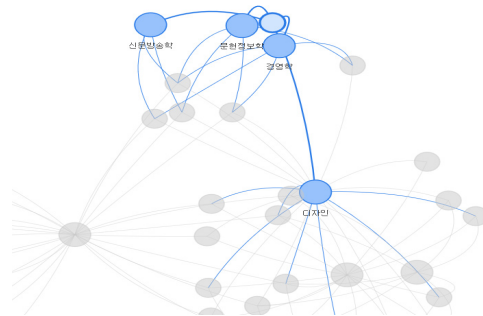
로 크게 3가지 유형, 즉 전문형, 일반형, 확대형으로 구분할 수 있다. 전문형은 속하는 주제 분야는 다른 주제 분야와 연관관계가 극히 제한적이며 한정된 주제 분야와 연관관계를 가지고 있다. 이 유형은 군집 1에 포함된 전문적인 주제 분야와 군집 2의 외곽에 위치한 주제 분야에서 나타나며, 대표적인 사례로 문헌정보학, 신문방송학, 경영학, 산업공학, 통계학이 있다. ‘문헌정보학’의 예를 들어보면, 극히 제한적인 연관관계를 가지고 있는데(〈그림 5〉 참조), 중간의 연결 노드

를 거쳐 ‘디자인’에 여러 갈래로 연결되어 있다(〈그림 5〉 왼쪽 참조). 또한 문헌정보학 주제 분야에서 중간의 연결 노드는 {문헌정보학, 신문방송학, 경영학, 디자인}이 서로 연관관계를 맺고 있었다(〈그림 5〉 오른쪽 참조). ‘신문방송학’을 살펴보면 ‘문헌정보학’과 유사하게 중간의 연결 노드를 통하여, {문헌정보학, 경영학, 디자인, 기타인문학}으로 연결된다(〈그림 6〉 참조). 이 유형은 ‘경영학’에서도 동일하게 나타난다.

둘째 유형은 ‘일반형’으로 일반적인 주제 분야

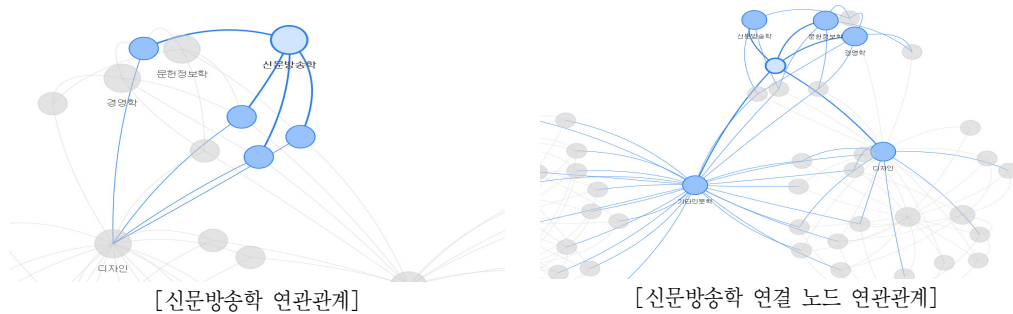


[문헌정보학 연관관계]



[문헌정보학 연결 노드 연관관계]

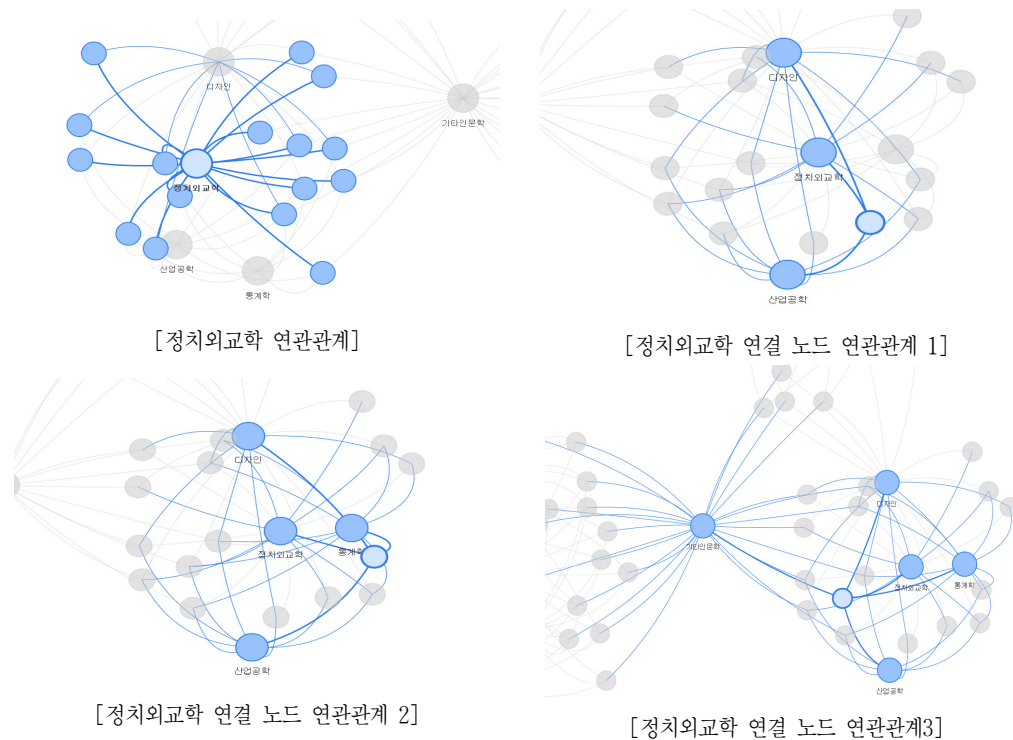
〈그림 5〉 문헌정보학 연관관계 사례



〈그림 6〉 신문방송학 연관관계 사례

가 가지고 있는 연관관계 유형으로 다양한 주제 분야와 연관관계를 맺고 있다. 이 유형은 군집 2의 중심에 있는 주제 분야와 군집 3의 외곽에 위치한 주제 분야에서 나타난다. 대표적인 사례로, 군집 2의 정치외교학, 군집 3의 무역학, 관

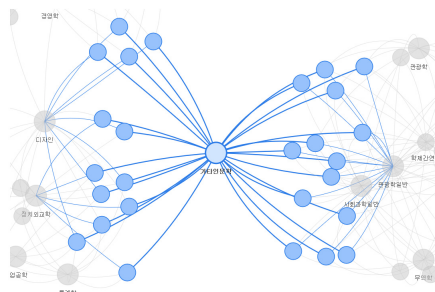
광학, 사회과학, 학제간연구가 있다. 정치외교학을 예로 들면, {정치외교학, 디자인, 산업공학}, {정치외교학, 디자인, 산업공학, 통계학}, {정치외교학, 디자인, 산업공학, 통계학, 기타인문학} 등 다양한 연관관계를 가지고 있다(〈그림 7〉 참조).



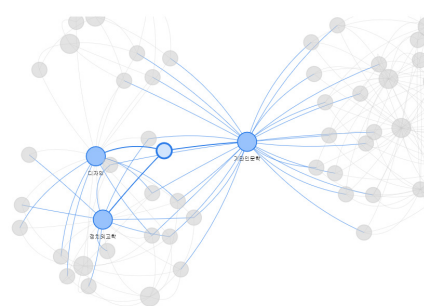
〈그림 7〉 정치외교학 연관관계 사례

셋째 유형은 ‘확대형’으로 ‘일반형’보다 더 다양하게 주제 분야 연관관계를 가지고 있으며, 군집을 연결하는 주제 분야와 군집 3과 같이 많은 주제 분야를 포함하는 군집의 중심에 위치한 주제 분야가 포함된다. 대표적인 사례로 디자인, 기타인문학, 사회과학일반, 관광학일반이 있다. 군집들의 연결고리 역할을 하는 ‘기타인문학’의 예를 들면, {기타인문학, 디자인, 정치

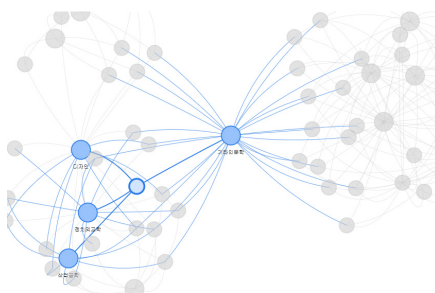
외교학}, {기타인문학, 디자인, 정치외교학, 산업공학}, {기타인문학, 사회과학일반, 무역학, 관광학일반, 사회과학}, {기타인문학, 사회과학일반, 관광학일반, 사회과학, 학제간연구}, {기타인문학, 관광학, 사회과학일반, 무역학, 사회과학, 관광학일반} 등 매우 다양한 형태의 연관관계를 보여주고 있다(〈그림 8〉 참조). 이 유형에 포함된 주제 분야는 확실한 전문적인 영역을 나타내지



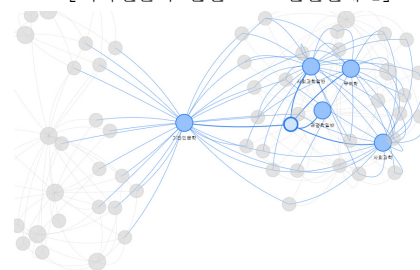
[기타인문학 연관관계]



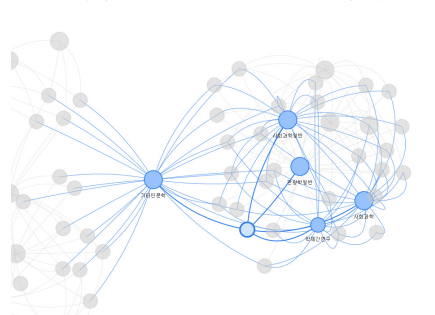
[기타인문학 연결 노드 연관관계 1]



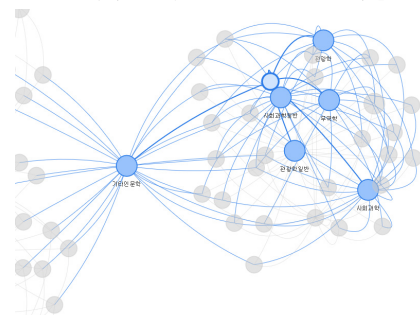
[기타인문학 연결 노드 연관관계 2]



[기타인문학 연결 노드 연관관계 3]



[기타인문학 연결 노드 연관관계 4]



[기타인문학 연결 노드 연관관계 5]

〈그림 8〉 기타인문학 연관관계 사례

못하고 일반적인 내용을 많이 포함하는 주제 분야이기 때문에, 관련된 주제 분야는 계속적으로 연관관계를 확대해 나갈 수 있을 것으로 보인다.

3.3 활용 및 제한점

빅데이터 관련 논문들을 조사한 결과, 빅데이터 연구는 다양한 주제 분야에서 수행되고 있으며, 많은 연구가 여러 주제 분야의 논문들을 인용하고 있음이 파악되었다. 주제 분야의 연관관계 분석 결과를 통해서 몇 가지 활용사항을 도출할 수 있다.

첫째, 빅데이터 연구자들은 타 주제 분야의 연구를 인용하여 연구를 수행하고 있어서, 학술정보 검색을 위한 새로운 주제 디렉토리를 구축할 필요가 있으며, 본 연구 결과를 주제 디렉토리 구축에 활용할 수 있다. 연구 결과에서도 보여주듯이 빅데이터 연구에서 ‘정치외교학’ 분야의 연구 논문을 인용하는 연구자는 ‘디자인’, ‘통계학’, 혹은 ‘산업공학’ 분야의 빅데이터 연구 논문도 인용할 가능성이 높기 때문에, 이들 주제 분야를 연계해서 학술정보서비스를 준비해야 할 것이다. 다시 말하면, 특정 연구자가 정치외교학 분야의 빅데이터 논문을 인용한다면, 디자인, 통계학, 산업공학 분야의 빅데이터 연구 정보를 제공해 줄 필요가 있다는 것이다.

둘째, 빅데이터를 연구하는 학문 분야에 따라 연관관계를 맺는 주제 분야의 범위가 다양하여, 학문 분야별로 다른 주제 디렉토리를 구축할 필요가 있다. 주제 분야 연관관계 범위는 전문형, 일반형, 확장형으로 구분되며, 유형에 따라 연관관계를 맺는 주제 분야 범위가 달라

진다. 문헌정보학이나 신문방송학과 같은 전문형 유형에서는 해당 주제 분야 및 인접 주제 분야 내에서 정보를 공유하는 형태를 보여주고 있다. 그러므로 ‘문헌정보학, 신문방송학, 경영학, 디자인’ 분야의 상호 연계관계를 보여주는 주제 디렉토리가 필요하다. 일반형 유형의 예는 ‘정치외교학, 디자인, 산업공학, 통계학, 기타인문학’이 있으며, 확대형 유형은 일반형에 비해 더 다양한 연관관계를 맺고 있는데, 예를 들면 ‘기타인문학, 사회과학일반, 무역학, 관광학일반, 사회과학’이 있고, ‘기타인문학, 관광학, 사회과학일반, 무역학, 사회과학, 관광학일반’처럼 더 확대된 연관관계가 존재한다. 그러므로 확대형 유형에서는 인용하는 주제 분야가 더 확대될 수 있다는 것을 염두에 두어야 할 것이다.

셋째, 빅데이터 분야 외에도 여러 학문 분야에서 융합적으로 연구하는 주제 분야에 대한 학술정보서비스는 해당 주제 분야를 연구하는 연구자들의 인용 행태를 분석하여 활용할 필요가 있다. 과거와 달리 오늘날에는 연구자들의 웹 DB 접근이 일반화되면서 연구자 스스로 적극적인 정보검색을 통하여 다양한 학술정보에 접근할 수 있다. 이용자에 대한 적극적인 서비스를 위해 도서관에서는 웹 DB 구독에 있어서 주제 분야 연관관계를 고려하여 상호 관계를 맺고 있는 주제 분야의 DB를 구독할 필요가 있다.

본 연구 결과의 제한점을 살펴보면 빅데이터 연구의 주제 분야를 객관적으로 판단하기 위해, KCI에서 학술지에 부여한 주제 분야를 활용하였다. 기존의 연구에서는 동일 주제 분야 내에서 분석하였기 때문에 논문 제목, 키워드 등에서 주제를 추출하였다. 이와 달리, 본 연구는 여러 주제 분야에서 빅데이터를 연구한 논문을

분석의 대상으로 삼았기 때문에, 논문 제목이나 키워드 대신, 논문이 게재된 학술지의 주제 분야를 분석의 대상으로 삼았다. 하지만, KCI의 학술지 주제 분야는 ‘기타인문학’, ‘사회학일반’, ‘공학일반’, ‘학제간연구’ 등 명확하지 않은 주제 분야가 있어, 연관관계 파악에 어려움이 있다.

또한, 연구자들이 생각하는 주제 분야와 학술지의 주제 분야가 일치하지 않을 수도 있어서, 주제 분야의 연관관계가 실제와 다를 수도 있다. 동시에 본 연구는 빅데이터 분야에서 인용 횟수가 높은 논문을 수록한 학술지의 주제 분야를 분석의 대상으로 삼았기 때문에, 연구 결과 해석에 있어서 한계가 있을 수 있다. 또한 동시인용이 1회 이루어진 경우도 분석에 포함되었기 때문에 일부 결과가 다를 수도 있으므로, 추후에는 데이터 수집 범위를 확장하여 더 많은 데이터를 활용하여 주제 분야사이의 연관 관계를 분석할 필요가 있다.

4. 결론 및 제언

본 연구의 목적은 빅데이터 연구가 이루어진 주제 분야 간에 연관관계를 분석하는데 있다. 연관관계는 빅데이터 연구자가 특정 주제 분야의 빅데이터 연구를 인용하면서 동시에 다른 주제 분야의 빅데이터 연구를 인용하는 경향을 보여주는 네트워크를 지칭한다. 연구 결과 빅데이터 연구와 관련된 22개 주제 분야가 추출

되었으며, 이들 주제 분야는 3가지 군집으로 구분되었다. 이들 군집 사이는 서로 연결하는 주제 분야가 존재하였는데, 군집 1과 군집 2 사이는 ‘디자인’이, 군집 2와 군집 3 사이는 ‘기타인문학’이 위치해 있었다. 주제 분야의 연관관계 유형을 분석한 결과, 연관관계의 복잡성에 따라 ‘전문형’, ‘일반형’, ‘확대형’으로 구분되었다. 전문형에는 문헌정보학, 신문방송학 등이 포함되었고, 일반형에는 정치외교학, 무역학, 관광학 등이 포함되었고, 확대형에는 기타인문학, 사회과학일반, 관광학일반 등이 포함되었다. 활용에 있어서 예를 들면, 빅데이터를 연구하는 연구자가 정치외교학 논문을 인용할 경우, 디자인, 통계학, 혹은 산업공학 주제 분야의 빅데이터 논문을 인용할 확률이 높으므로, 도서관에서는 이러한 경향을 고려하여 학술정보서비스를 준비해야 한다.

제언으로 첫째, 연구 논문의 주제 분야를 KCI에서 제공하는 학술지의 주제 분야를 사용하였지만, 이는 연구자가 생각하는 주제 분야와 다를 수가 있어, 다른 융합적인 주제 분야를 대상으로 데이터를 분석할 필요가 있을 것이다. 둘째, 국외의 빅데이터 연구를 대상으로 한 조사에서 나타난 관련된 주제 분야와 본 연구의 결과와 차이가 있는데, 그 원인을 분석할 필요가 있다. 셋째, 본 연구는 분석할 주제 분야 선택을 위해 동시 인용분석을 적용하였지만, 역으로 서지 결합법으로 주제 분야를 선택한 후, 주제 분야 연관관계를 분석할 필요가 있다.

참 고 문 헌

- 고재창, 조근태, 조윤희 (2013). 키워드 네트워크 분석을 통해 살펴본 기술경영의 최근 연구동향. *지능정보연구*, 19(2), 101-123.
- 곽선영, 정은경 (2012). 복수저자기반 동시인용분석을 활용한 지적구조 분석: 경제학을 중심으로. *정보관리학회지*, 29(1), 115-134. <http://dx.doi.org/10.3743/KOSIM.2012.29.1.115>
- 김용 (2012). Apriori 알고리즘 기반의 개인화 정보 추천시스템 설계 및 구현에 관한 연구. *한국비블리아학회지*, 23(4), 283-308.
- 김희정 (2011). 네트워크 분석을 기반으로 한 웹 아카이빙 주제영역 연구. *한국비블리아학회지*, 22(2), 235-248.
- 박옥남 (2011). 정보조직 지식구조에 대한 연구: 2000년~2011년 학술논문을 중심으로. *한국비블리아학회지*, 22(3), 247-267.
- 백봉현, 하일규 (2017). XML 구조를 이용한 공공 빅데이터의 선별 저장 및 시각화 방법. *한국정보통신학회논문지*, 21(12), 2305-2311. <https://doi.org/10.6109/jkiice.2017.21.12.2305>
- 손혜인, 남영준 (2016). 기록관리학 분야 국내 학술지의 연구동향에 관한 연구: 『한국기록관리학회지』와 『기록학연구』를 중심으로. *정보관리학회지*, 33(1), 85-110.
<http://dx.doi.org/10.3743/KOSIM.2016.33.1.085>
- 이기현, 정효정, 송민 (2015). 문헌정보학 분야 핵심 학술지들의 가중 주제-방법 네트워크 분석. *한국문헌정보학회지*, 49(3), 457-488. <http://dx.doi.org/10.4275/KSLIS.2015.49.3.457>
- 이보람, 정은경 (2016). 학술지 단위 서지결합분석을 통한 빅데이터 연구분야의 학제적 구조에 관한 연구. *정보관리학회지*, 33(3), 133-154. <http://dx.doi.org/10.3743/KOSIM.2016.33.3.133>
- 이수상 (2013). 연구자 네트워크의 중심성과 연구성과의 연관성 분석: 국내 기록관리학 분야 학술논문을 중심으로. *한국도서관·정보학회지*, 44(3), 405-428.
- 이재윤 (2015). 문헌동시인용 분석을 통한 한국 문헌정보학의 연구 전선 파악. *정보관리학회지*, 32(4), 77-106. <http://dx.doi.org/10.3743/KOSIM.2015.32.4.077>
- 이정미 (2013). 빅데이터의 이해와 도서관 정보서비스에의 활용. *한국비블리아학회지*, 24(4), 53-73. <http://dx.doi.org/10.14699/kbiblia.2013.24.4.053>
- 이환수, 임동원, 조항정 (2013). 빅데이터 시대의 개인정보 과잉이 사용자 저항에 미치는 영향. *지능정보연구*, 19(1), 125-139.
- 장임숙, 장덕현, 이수상 (2011). 다문화연구의 지식구조에 관한 네트워크 분석. *한국도서관·정보학회지*, 42(4), 353-374.
- 정동열 (2016). 한국문헌정보학 학문영역의 변천 과정에 관한 고찰: 학술지 논문 주제분석을 기반으로.

- 한국문헌정보학회지, 50(4), 5-34. <http://dx.doi.org/10.4275/KSLIS.2016.50.4.005>
- 조수권, 김성범 (2012). 텍스트마이닝을 활용한 산업공학 학술지의 논문 주제어간 연관관계 연구. 대한 산업공학학회지, 38(1), 67-73.
- 조재인 (2011). 네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석. 정보관리학회지, 28(4), 65-83. <http://dx.doi.org/10.3743/KOSIM.2011.28.4.065>
- 최이랑 (2015). 국내 기록관리학 연구동향에 관한 연구: 최근 10년간(2004-2013) 학술논문을 중심으로. 기록학연구, 43, 147-177.
- Mishra, D., Luo, Z., Jiang, S., Papadopoulos, T., & Dubey, R. (2017). A bibliographic study on big data: Concepts, trends and challenges. Business Process Management Journal, 23(3), 555-573.
- Peng, T., Zhang, L., Zhong, Z., & Zhu, J. J. (2012). Mapping the landscape of internet studies: Text mining of social science journal articles 2000-2009. New Media & Society, 15(5), 644-664.
- Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996-2005: Introducing author bibliographic-coupling analysis. Journal of the American Society for Information Science and Technology, 59(13), 2070-2086.
- Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006-2010: An author cocitation and bibliographic coupling analysis. Journal of the Association for Information Science and Technology, 65(5), 995-1006.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Back, Bong Hyun, & Ha, Il-Kyu (2017). A method for selective storing and visualization of public big data using XML structure. Journal of the Korea Institute of Information and Communication Engineering, 21(12), 2305-2311. <https://doi.org/10.6109/jkiice.2017.21.12.2305>
- Cho, Jane (2011). A study for research area of library and information science by network text analysis. Journal of the Korean Society for Information Management, 28(4), 65-83. <http://dx.doi.org/10.3743/KOSIM.2011.28.4.065>
- Cho, Su Gon, & Kim, Seoung Bum (2012). Finding meaningful pattern of key words in IIE transactions using text mining. Journal of the Korean Institute of Industrial Engineers, 38(1), 67-73.
- Choi, Yilang (2015). A study on the research trends of archival studies in Korea: Focused on

- research papers between 2004 and 2013. *The Korean Journal of Archival Studies*, 43, 147-177.
- Jang, Im-Sook, Chang, Durk-Hyun, & Lee, Soo-Sang (2011). The knowledge structure of multicultural research papers in Korea. *Journal of Korean Library and Information Science Society*, 42(4), 353-374.
- Jeong, Dong Youl (2016). A study on the historical process of library and information science domain in Korea: Based on the subject analysis of academic journal. *Journal of the Korean Society for Library and Information Science*, 50(4), 5-34.
<http://dx.doi.org/10.4275/KSLIS.2016.50.4.005>
- Kho, Jaechang, Cho, Kuentae, & Cho, Yoonho (2013). A study on recent research trend in management of technology using keywords network analysis. *Journal of Intelligence and Information Systems*, 19(2), 101-123.
- Kim, Heejung (2011). A study on web archiving subject analysis based on network analysis. *Journal of the Korean Biblia Society for Library and Information Science*, 22(2), 235-248.
- Kim, Yong (2012). A study on design and implementation of personalized information recommendation system based on Apriori algorithm. *Journal of the Korean Biblia Society for Library and Information Science*, 23(4), 283-308.
- Kwak, Sun-Young, & Chung, EunKyung (2012). Domain analysis on economics by utilizing cocitation analysis of multiple authorship. *Journal of the Korean Society for Information Management*, 29(1), 115-134. <http://dx.doi.org/10.3743/KOSIM.2012.29.1.115>
- Lee, Boram, & Chung, Eunkyung (2016). A study on interdisciplinary structure of big data research with journal-level bibliographic-coupling analysis. *Journal of the Korean Society for Information Management*, 33(3), 133-154. <http://dx.doi.org/10.3743/KOSIM.2016.33.3.133>
- Lee, Hwansoo, Lim, Dongwon, & Zo, Hangjung (2013). Personal information overload and user resistance in the big data age. *Journal of Intelligence and Information System*, 19(1), 125-139.
- Lee, Jae Yun (2015). Identifying the research fronts in Korean library and information science by document co-citation analysis. *Journal of the Korean Society for Information Management*, 32(4), 77-106. <http://dx.doi.org/10.3743/KOSIM.2015.32.4.077>
- Lee, Jeong-Mee (2013). Understanding big data and utilizing its analysis into library and information services. *Journal of the Korean Biblia Society for Library and Information Science*, 24(4), 53-73. <http://dx.doi.org/10.14699/kbiblia.2013.24.4.053>
- Lee, Keehoen, Jung, Hyojung, & Song, Min (2015). Weighted subject-method network analysis of library and information science studies. *Journal of the Korean Society for Library and*

- Information Science, 49(3), 457-488. <http://dx.doi.org/10.4275/KSLIS.2015.49.3.457>
- Lee, Soo-Sang (2013). Analytical study on the relationship between centralities of research networks and research performances. *Journal of Korean Library and Information Science Society*, 44(3), 405-428.
- Park, Ok Nam (2011). Knowledge structures in knowledge organization research: 2000-2011. *Journal of the Korean Bibliography Society for Library and Information Science*, 22(3), 247-267.
- Sohn, Hye In, & Nam, Young Joon (2016). A study on the research trends of archives management in Korea: Focused on the *Journal of Records Management & Archives Society of Korea* and the *Korean Journal of Archival Studies*. *Journal of the Korean Society for Information Management*, 33(1), 85-110. <http://dx.doi.org/10.3743/KOSIM.2016.33.1.085>

[부록 1] 분석 대상 주제 분야 매트릭스

	경영학	관광학	관광학 일반	기초법	기타공학 일반	기타과학기술학	기타법학	기타인문학	디자인	무역학	문헌정보학	법학	사회과학	사회과학 일반	사회과학 정보	산업공학	신문방송학	예술 일반	정보통신이론	정치외교학	통계학	학제간연구	합계
1				1								1											2
2				1								1		1	1		1						5
3								1						1	1	1	1						5
4														1			1						2
5														1		1							2
6																			1			1	2
7		1						1		1			1										4
8						1	1															1	3
9		1				1																	2
10		1								1												1	3
11		1			1					1						1							4
12		1																				1	2
13		1						1									1						3
14											1					1							2
15				1			1					1											3
16												1				1							2
17												1	1										2
18				1								1											2
19				1			1					1											3
20												1	1										2
21				1			1					1											3
22					1											1							2
23					1													1					2
24																1						1	2
25						1						1									1		3
26	1													1			1					1	4
27												1				1							2
28						1						1											2
29																1						1	2
30								1											1				2
31	1					1																	2
32	1			1								1		1		1							5
33													1				1						2
34							1					1											2
35																		1				1	2
36					1										1								2
37	1																				1		2
38														1			1						2
39						1										1						1	2
40															1	1			1				3
41				1								1											2
42																1			1				2
43																	1				1		2
44								1													1		2
45				1								1				1	1						4

	경영학	관광학	관광학일반	기초법	기타공학일반	기타과학기술학	기타법학	기타인문학	디자인	무역학	문헌정보학	법학	사회과학	사회과학일반	사회과학정보	산업공학	신문방송학	예술일반	정보통신이론	정치외교학	통계학	학제간연구	합계
46				1			1										1						3
47						1						1			1								3
48		1						1								1							3
49				1								1				1							3
50								1		1						1					1		4
51	1																				1		2
52															1	1							2
53	1		1																				2
54											1											1	2
55		1								1													2
56		1									1												2
57		1												1									2
58	1							1								1						1	4
59		1	1					1					1										4
60														1			1						2
61									1		1												2
62															1	1			1				3
63							1					1											2
64												1				1							2
65												1	1										2
66																					1	1	2
67						1									1							1	3
68						1									1							1	3
69				1			1					1											3
70							1					1											2
71												1			1								2
72	1	1																					2
73														1		1							2
74		1						1		1												1	4
75								1								1							2
76		1						1								1							3
77				1			1					1			1							1	5
78														1			1						2
79											1					1							2
80								1						1			1						3
81	1			1																			2
82								1							1	1						1	4
83															1		1				1		3
84											1				1	1	1					1	5
85						1		1							1		1						4
86							1					1											2
87				1								1			1								3
88																					1	1	2
89						1												1					2
90															1	1						1	3
91		1	1					1															3
92																1						1	2
93		1								1													2

	경영학	관광학	관광학일반	기초법	기타공학일반	기타과학기술학	기타법학	기타인문학	디자인	무역학	문헌정보학	법학	사회과학	사회과학일반	사회과학정보	산업공학	신문방송학	예술일반	정보통신이론	정치외교학	통계학	학제간연구	합계
94		1														1							2
95		1	1					1		1			1	1								1	7
96					1											1							2
97					1	1																	2
98						1																1	2
99													1									1	2
100		1												1									2
101								1			1												2
102								1	1							1				1	1		5
103		1																			1		2
104						1																1	2
105																1					1		2
106				1			1									1							3
107																					1	1	2
108														1		1	1						3
109							1					1											2
110	1							1	1		1						1						5
111		1				1																	2
112		1								1													2
113											1				1	1							3
114						1		1															2
115																			1		1		2
116											1					1	1						3
117					1													1					2
118		1								1				1									3
119														1		1					1		3
120					1											1							2
121					1											1							2
122																1						1	2
123		1																			1		2
124					1											1					1		3
125		1														1							2
126																					1	1	2
127																1					1		2
128					1						1		1			1							4
129						1																1	2
130								1											1				2
131							1					1		1		1	1						5
132		1													1								2
133						1												1					2
134		1	1																				2
	10	27	5	16	11	18	14	22	3	10	11	29	8	19	19	45	20	5	7	1	19	29	348